NPSCS-94-008

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Efficient Caption-based Retrieval of Multimedia Information

by Neil C. Rowe

October 1993

Prepared for:

DARPA
3701 N. Fairfax Drive
Arlington, VA 22203-1714

Naval Postgraduate School
Monterey, CA 93943-5118

NAVAL POSTGRADUATE SCHOOL

Monterey, California

REAR ADMIRAL T. A. MERCER
Superintendent

HARRISON SHULL
Provost

This work was sponsored by DARPA as part of the 13 Project under AO 8939, and by the Naval Postgraduate School under funds provided by the Chief for Naval Operations

This report was prepared by:

Yutaka Kanayama
Associate Chairman for
Technical Research

PAUL J. MARTO
Dean of Research

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPSCS-94-008	5. MONITORING ORGANIZATION REPORT NUMBER(S) Naval Postgraduate School
6a. NAME OF PERFORMING ORGANIZATION Computer Science Dept. Naval Postgraduate School 6b. OFFICE SYMBOL *(if applicable)* CS	7a. NAME OF MONITORING ORGANIZATION DARPA
6c. ADDRESS *(City, State, and ZIP Code)* Monterey, CA 93943	7b. ADDRESS *(City, State, and ZIP Code)* 3701 N. Fairfax Drive Arlington, VA 22203-1714
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Postgraduate School 8b. OFFICE SYMBOL *(if applicable)* NPS	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AO 3839
8c. ADDRESS *(City, State, and ZIP Code)* Monterey, CA 93943	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE *(Include Security Classification)*
Efficient Caption-based Retrieval of Multimedia Information

12. PERSONAL AUTHOR(S)
Neil C. Rowe

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Progress	FROM 10/92 TO 9/93	October 1993	9

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	multimedia, captions, retrieval, databases, pausing

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

We describe MARIE-1 and MARIE-2, information retrieval systems for multimedia data. They exploit captions on the data and perform natural-language processing of them and English retrieval requests. Some content analysis of the data is also performed to obtain additional descriptive information. The key to getting this approach to work is sufficiently-fast processing. We achieve this by decomposing the problem into "information filters" and applying a new theory of optimal information filtering which we have developed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dennis M. Volpano	22b. TELEPHONE *(Include Area Code)* (408) 656-3091 22c. OFFICE SYMBOL CSVo

Efficient Caption-based Retrieval of Multimedia Information

Neil C. Rowe[1]

Code CS/Rp, Department of Computer Science

Naval Postgraduate School

Monterey, CA USA 93943

rowe@cs.nps.navy.mil

ABSTRACT

We describe MARIE-1 and MARIE-2, information retrieval systems for multimedia data. They exploit captions on the data and perform natural-language processing of them and English retrieval requests. Some content analysis of the data is also performed to obtain additional descriptive information. The key to getting this approach to work is sufficiently-fast processing. We achieve this by decomposing the problem into "information filters" and applying a new theory of optimal information filtering which we have developed.

[1] This work was sponsored by DARPA as part of the I3 Project under AO 8939, and by the U. S. Naval Postgraduate School under funds provided by the Chief for Naval Operations. Discussions with Amr Zaky improved this paper.

1. Introduction

The MARIE project has been investigating information retrieval of multimedia data using a new idea: putting primary emphasis on caption processing. Even though content analysis methods such as substring searching for text media and shape matching for picture media can obviate captions, content analysis usually requires unacceptably-large amounts of time at retrieval time. Captions can be though of as cachings of the results of content analysis, created either manually by a user describing a multimedia datum, automatically by computerized content analysis, or some combination of both; but they can also include auxiliary information like the date or customer for a photograph. Since captions can be considerably smaller than the media data they describe, checking captions before retrieving media data can save time if it can rule out many bad matches quickly. In other words, caption information can be passed through fast "information filters" [1] to rule out media data irrelevant to a user needs.

However, caption processing does not necessarily give faster multimedia retrieval. The terms of the caption are perhaps synonyms or subterms of those supplied by a user during retrieval, in which case a complete thesaurus of synonyms and a complete type hierarchy covering more general and specialized terms should be available for use when matching the caption during information retrieval [21]. Furthermore, to obtain high query recall and precision, user-supplied captions should be subject to natural-language processing to determine the correct word senses and how the words relate, to get beyond the limits of keyword matching on the caption [11]. These additional processing needs can make caption processing slow. So the MARIE project is concerned with methods of improving efficiency of caption-based approach to information retrieval. This paper reports on three important directions that we have explored recently: an efficient statistical parser for natural language, special content-analysis methods, and using sampled parameters to find the optimal execution strategy for retrievals.

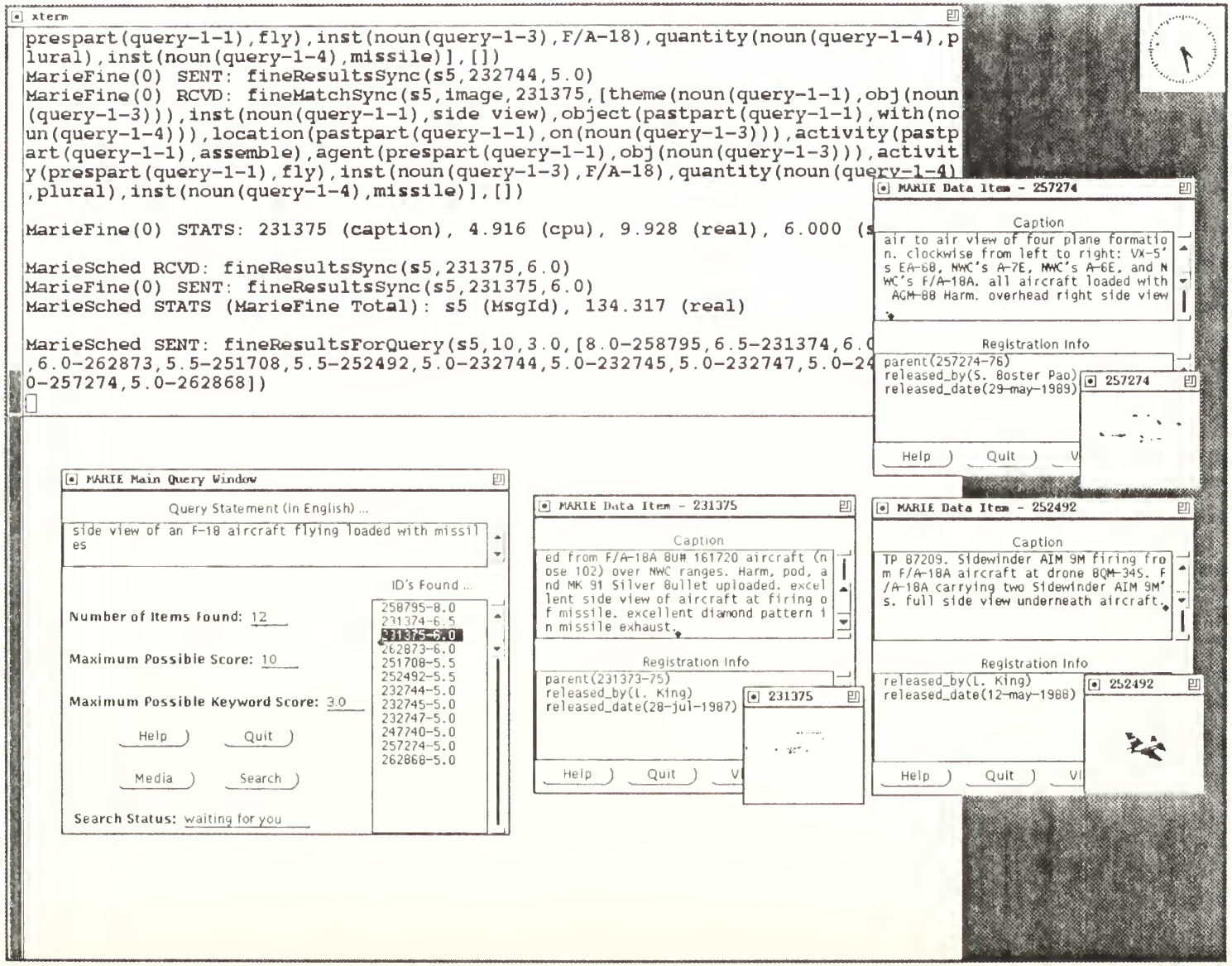
While the MARIE project is intended for multimedia information retrieval in general, we have used as testbed the Photo Lab of the Naval Air Warfare Center (NAWC-WD), China Lake California USA. This is a library of approximately 100,000 pictures and 37,000 captions for those pictures. The pictures cover all activities of the center, including pictures of equipment, tests of equipment, administrative documentation, site visits, and public relations. With so many pictures, many of which look virtually identical, captions are indispensable to find what

a user is looking for. But the existing computerized keyword system for finding pictures from their captions is unhelpful, and is mostly ignored by personnel. [17] reports on MARIE-1, a prototype implementation that we developed for them, a system that appears much more in the direction of what users want. Figure 1 shows an example retrieval from MARIE-1, for the query "side view of an F-18 aircraft flying loaded with missiles"; 12 pictures were found (with fits ranging from 5.0 to 8.0), three of which are displayed in the bottom right with their associated registration information, and the top of the upper left box shows the semantic interpretation constructed for the query. MARIE-1 took a man-year to construct and only handled 220 pictures from the database. To handle the full database, efficiency and implementation-difficulty concerns become paramount. MARIE-2, currently under development, will address these

2. Statistical natural-language parsing

Some natural-language processing beyond keyword matching seems important for visual and audio multimedia because relationships between components are more important for them than for most documents. For instance, users should demand that "tank target" should not match just any caption mentioning "tank" and "target", nor "steel airplane propeller" match a caption mentioning "steel", "airplane", and "propeller" separately, nor "missile on dirt" match "dirt on missile". Similarly, users should expect a type and part-whole hierarchies to be used, so "closeup of wing markings" should match "view of wing". To permit such reasonable behavior, we will need to do parsing and some semantic interpretation of each caption and query.

MARIE-1 uses the standard approach of intelligent natural-language processing for information retrieval [9, 13, 19] of hand-coding of lexical and semantic information for the words in a narrow domain. This approach would be laborious and near-unworkable for the 32,000 distinct words in the 100,000-caption NAWC database. But a new approach to natural-language processing has emerged in the last few years, statistical parsing. It assigns probabilities of co-occurrence to sets of words, and uses these probabilities to guess the most likely interpretation of a sentence. The probabilities can be derived from statistics on a corpus, a representative set of example sentences, and they can capture fine semantic distinctions that would otherwise require additional lexicon information. Statistical parsing seems an excellent way to implement MARIE-2 since it replaces invocation of many


xterm
prespart(query-1-1),fly),inst(noun(query-1-3),F/A-18),quantity(noun(query-1-4),plural),inst(noun(query-1-4),missile)],[])
MarieFine(0) SENT: fineResultsSync(s5,232744,5.0)
MarieFine(0) RCVD: fineMatchSync(s5,image,231375,[theme(noun(query-1-1),obj(noun(query-1-3))),inst(noun(query-1-1),side view),object(pastpart(query-1-1),with(noun(query-1-4))),location(pastpart(query-1-1),on(noun(query-1-3))),activity(pastpart(query-1-1),assemble),agent(prespart(query-1-1),obj(noun(query-1-3))),activity(prespart(query-1-1),fly),inst(noun(query-1-3),F/A-18),quantity(noun(query-1-4),plural),inst(noun(query-1-4),missile)],[])
MarieFine(0) STATS: 231375 (caption), 4.916 (cpu), 9.928 (real), 6.000 (s
MarieSched RCVD: fineResultsSync(s5,231375,6.0)
MarieFine(0) SENT: fineResultsSync(s5,231375,6.0)
MarieSched STATS (MarieFine Total): s5 (MsgId), 134.317 (real)
MarieSched SENT: fineResultsForQuery(s5,10,3.0,[8.0-258795,6.5-231374,6.0,6.0-262873,5.5-251708,5.5-252492,5.0-232744,5.0-232745,5.0-232747,5.0-247740,5.0-257274,5.0-262868])
MARIE Data Item - 257274
Caption
air to air view of four plane formation. clockwise from left to right: VX-5's EA-6B, NWC's A-7E, NWC's A-6E, and NWC's F/A-18A. all aircraft loaded with AGM-88 Harm. overhead right side view
Registration Info
parent(257274-76)
released_by(S. Boster Pao)
released_date(29-may-1989)
Help
Quit
257274
MARIE Main Query Window
Query Statement (in English) ...
side view of an F-18 aircraft flying loaded with missiles
ID's Found ...
Number of Items Found: 12
Maximum Possible Score: 10
Maximum Possible Keyword Score: 3.0
Help
Quit
Media
Search
Search Status: waiting for you
258795-8.0
231374-6.5
231375-6.0
262873-6.0
251708-5.5
252492-5.5
232744-5.0
232745-5.0
232747-5.0
247740-5.0
257274-5.0
262868-5.0
MARIE Data Item - 231375
Caption
ed from F/A-18A BUH 161720 aircraft (nose 102) over NWC ranges. Harm, pod, and MK 91 Silver Bullet uploaded. excellent side view of aircraft at firing of missile. excellent diamond pattern in missile exhaust.
Registration Info
parent(231373-75)
released_by(L. King)
released_date(28-jul-1987)
Help
Quit
231375
MARIE Data Item - 252492
Caption
TP 87209. Sidewinder AIM 9M firing from F/A-18A aircraft at drone BQM-34S. F/A-18A carrying two Sidewinder AIM 9M's. full side view underneath aircraft.
Registration Info
released_by(L. King)
released_date(12-may-1988)
Help
Quit
252492

laboriously-handcrafted semantic routines with a few simple and fast calculations on statistics automatically acquired from a corpus with many similar sentences.

Statistical parsing is especially well suited for information-retrieval applications because they already have a statistical aspect: They find data that is probable, but not guaranteed, to satisfy a user. Also, good information retrieval does not require the full natural-language understanding that hand-tailored semantic routines provide: Understanding of the words involved in matching is not generally helpful for matching beyond the synonym and hierarchical type and part information for those words. For instance, the query "missile mounted on aircraft" should match all three of "Sidewinder on F-18", "Sidewinder attached to wing pylon", and "Pylon mounted AIM-9M Sidewinders" since "Sidewinder" and "AIM-9M" are types of missiles, "F-18" as a kind of aircraft, and "on" and "attached" mean the same thing as "mounted". In fact, the MARIE-1 captions were often very imprecise with verbs, so that detailed semantic analysis of verbs and their cases in captions was unhelpful. Parsing is still essential to connect related words in a caption, so to recognize that the three examples above have the same deep semantic structure. But for information-retrieval applications, this parser can be simpler than one required for full natural-language understanding, with fewer grammatical categories and fewer rules.

Creating the full synonym list, type hierarchy, and part hierarchy for applications of the size of the NAWC-WD database (32,000 words) is some work. Fortunately, most of this job for any English application has been already accomplished in the Wordnet system [12] 1990), a large thesaurus system that includes this information plus rough word frequencies and morphological processing. We converted its information for the NAWC-WD words into a Prolog format compatible with the rest of MARIE-2, and used this as our lexicon for parsing and interpretation. So the basic meaning assigned to a noun or verb is that it is a subtype of the concept designed by its name in the type hierarchy, with additional pieces of meaning added by its relationships (like modification) to other words in the sentence. Wordnet also includes extensive lists of synonyms; using the rough word-frequency information, we designated the most common one of each synonym set as the "standard alias", and store only the type and part pointers for this word, which considerably shortens the lexicon.

2.1. Statistical parsing techniques

This approach can mean fast processing since we just append the type and relationship specifications for all the words in a sentence, resolving references using the parse tree, to obtain a "meaning list" or semantic graph, following the paradigm of [6]. But this can still be slow because we need to find all the reasonable interpretations of a sentence in order to rank them, and most sentences have multiple interpretations. To simplify matters, we restricted the grammar to binary parse rules (context-free grammar rules with only one or two symbols for the replacement). Then the likelihood of an interpretation can be found by assigning probabilities to word senses and grammar rules. If we could assume near-independence of the probabilities of each part of the sentence, we could multiply them to get the probability of the whole sentence [8]. This is mathematically equivalent to taking the sum of the logarithms of the probabilities, and hence a branch-and-bound search could be done to quickly find the N best parses of the a sentence.

But words of sentences are obviously not often independent or near-independent. Statistical parsing often exploits the probabilities of strings of successive words in a sentence [10]. However, with our binary parse rules, a simpler and more semantic approach is to only consider the probability of co-occurrence of the two subparses in the binary rule. For example, in parsing "F-18 landing" by the rule "NP -> NOUN PARTICIPLE", the probability assigned to this rule should reflect the likelihood of an F-18 in particular doing a landing in addition to the probability of using this rule. The co-occurrence probability for "F-18" and "land" is especially helpful because it is unexpectedly large, since there are only a few things in the world that land. Estimates of co-occurrence probabilities can inherit in the type hierarchy [14]. So if we have insufficient statistics in our corpus about how often an F-18 lands, we may have enough on how often an aircraft lands; and assuming that F-18s are typical of aircraft in this respect, we can estimate how often F-18s land. The second word can separately be generalized too, so we can use statistics on "F-18" and "moving", or both the words can be simultaneously generalized, so we can use statistics on "aircraft" and "moving". The objective should be to find some statistics that can be reliably used to estimate the co-occurrence probability of the words.

To keep this number of possible co-occurrence probabilities manageable, it is important to restrict them to two-probability. When parse rules recognize multiword sequences as grammatical units, those sequences can be

reduced to "headwords". For instance, "the big F-18 from China Lake landing at Armitage Field" can be parsed by the rule "NP => NP PARTP" and the same co-occurrence probability used, since "F-18" is the principal noun and hence headword of the noun phrase "the big F-18 from China Lake", and "landing" is the participle and hence headword of the participial phrase "landing at Armitage Field".

The statistical database for binary co-occurrence statistics will need careful design because the data will be sparse and there will be many small entries. For instance, for the NAWC-WD captions with 32,000 possible words and 9,000 superconcepts and aliases of those words, there are 26,000 distinct lexicon entries after equivalent aliases are removed and all word senses are included. This means 343 million possible co-occurrence pairs, but the total of all their counts can only be 605,000, the total number of word instances in all the captions. Our database uses four search trees indexed on the first word, the part of speech + word sense of the first word, the second word, and the part of speech + word sense of the second word; it stores the count for that word pair. It is important to store counts rather than probabilities to save storage and reduce work on update. Various compression techniques can further reduce the size of this database, but one in particular in especially useful, elimination of data that can be closely approximated from other counts [14] using sampling theory. For instance, if "F-18" occurs 10 times in the corpus, all kinds of aircraft occur 1000 times, and there are 230 occurrences of aircraft landing, estimate the number of "F-18 landing"s in the corpus as 230* 10/1000=2.3; if the actual count is within a standard deviation of the value, do not store it in the database. The standard deviation when n is the size of the subpopulation, N is the size of the population, and A the count for the population, is $\sqrt{A(N-A)(N-n)/nN^2(N-1)}$ [4]. Such calculations require also "unary" counts stored with each word or standard phrase, but there are far fewer of these. (While unary counts also directly affect the likelihood of a particular sentence, that effect can be ignored in judging different interpretations of a sentence since it is constant.)

3. Integrating content analysis

Another way to obtain descriptive caption information for a multimedia datum is to analyze its content directly, as in [2, 5]. For text data this can be parsing and summarization, but for pictures, audio, and video it is more complex. Audio can be reduced to a picture by a Fourier transform, and video can be converted into a sequence

of still pictures. Thus the central problem for content analysis of multimedia is one of recognizing and classifying shapes in a two-dimensional picture. For instance, aircraft in NAWC-WD photographs are usually the only objects with four bumps in two symmetrical pairs; even if the caption doesn't say so, such a shape should be considered evidence of an aircraft. We developed some powerful domain-independent picture processing methods in [18]; additional domain-dependent knowledge is also needed to classify shapes. Then qualitative relationships between the shapes can be determined. The shape and relationship facts can be collected as a visual summary of the picture, and this can be merged with explicit textual caption information.

Content analysis of pictures can be complex because interesting ones (or audio or video) can contain many different shapes and relationships between them. The work may be done when multimedia data are added to the databases, and different processors can work on different parts of the picture simultaneously to get results faster. To avoid creating unwieldy captions, the amount of such information can be limited to that for the highest-priority shapes (like aircraft for the NAWC-WD pictures, or the long sounds in [18]). Alternatively, we can store only information about regions mentioned in the caption, but this requires we relate the caption graph and content-analysis graph. In general, the caption graph, excluding nondepictable concepts like "view", "test", and dates, will be a subgraph of the content-analysis graph, and a subgraph isomorphism problem must be solved to merge the two into a single graph. The subgraph isomorphism problem is NP-complete in general, although this application of it provides are variety of special heuristics to exploit. But the resulting consensus graph will provide better picture-description information than either graph alone.

Just as captions have linguistic foci, pictures that depict have visual foci, something not true of pictures in general. That is, if a picture is to be considered a "good" depiction of something, and worth storing in a multimedia library, the object(s) depicted usually can be inferrable from the picture alone. However, photography is a less precise enterprise than entering captions because photographs sometimes must be taken in a hurry, and the best angle to the subject or best distance from the photographic subject is not always possible, and it is also much harder to "edit" the results. So visual focus can only be established by a set of factors that positively correlate with it.

We have identified six major factors that can be applied to the regions identified in a picture to rate how likely a

region or set of contiguous regions is to be a visual focus. First, a visual focus tends to be a big region or set of regions (with exceptions for photographs illustrating the context of some subject). Second, a visual focus tends to be surrounded be a strong edge, or clear discontinuity in brightness, color, or texture. Third, a visual focus tends to be either a uniform color or color mix, although its brightness may vary considerably. Fourth, a visual focus tends not to touch the boundary of picture, though large objects can touch a little (with one major exception: People and some animals are generally considered depicted if their faces are depicted.) Fifth, a visual focus has its center of mass close to the center of the photograph. Sixth, there are few other regions or region clusters having the same properties as the visual focus (with exceptions for some natural pictures like those of flowers in a field).

So early visual processing should be adjusted, in thresholds and in the techniques used, to find such a region or regions, using parameters for textural discrimination between regions if necessary; [18] describes the techniques we are exploring for this in one domain. The tendency of these six factors to correlate with visual focus naturally maps to a neural net with the factors as inputs. The neural net should be trainable, since there are no human experts to consult with on the proper weightings of the factors. The weights on the factors also need adjusting to the domain and picture type within the domain because they can obviously vary significantly. For example, for most NAWC-WD pictures, the fourth and fifth factors are very important, and the first factor is quite unimportant because there many occasions when the context in which a small object is embedded is more important than the object. But process documentation pictures, type (4) of the last section, are often taken in a hurry at NAWC-WD, and for them the first, fourth, and fifth factors must all be weighted lightly.

Another way to handle large captions derived from content analysis is to use supercaptions, captions describing common features of sets of captions. Explicit supercaptions occur frequently with the NAWC-WD pictures for sets of photographs taken of the same subject in the same picture-taking session. On querying, the supercaption can be matched first to the user query, and if it passes, the full caption can be matched. Supercaptions can form a hierarchy, possibly quite different from the type hierarchy. We have done some simple experiments using supercaptions, with positive results.

4. Finding an efficient execution plan for a query

One objection raised to natural-language processing for information retrieval [20] is that even if you get can the parsing and meaning-list construction to be done quickly, you still have other problems, including a different subgraph isomorphism problem, to solve in matching the query graph (or "meaning list") to candidate caption graphs. The latter took an average of two seconds per query-caption pair on a Sun-4 workstation using a simple algorithm in MARIE-1. Certainly the content-analysis methods of the last section can be slow. Furthermore, multimedia data can be large and will be usually slow to retrieve under traditional database methods. We believe that speed problems for multimedia retrieval be signifieantly minimized by appropriate prior use of "information filters", processes that rule out matches using simple polynomial-time criteria. We will assume here that information filters guarantee perfect recall although not necessarily perfect precision, or that they never rule out an acceptable data match. Signature matehing [7] is the most familiar information filter for multimedia retrieval, but it can be done more than once for an application [3], and filters based on semantic or "intelligent" criteria are also useful.

MARIE-1 got much power from "coarse-grain" filters that extracted nouns from the query and retrieved indexes of captions that mentioned those nouns or their superconcepts (their generalizations in the type hierarchy). In subsequent work, [15] reported significant power from a filter that assigns a set of possible categories to each picture based on its intended purpose, and matches these to categories inferred for the query. [16] then reported experiments with a "registration-data" filter to extract restrietions covered by the bookkeeping information for each picture, information that can be stored separately in a relational database; the filter executes SQL queries on this database, and rules out pictures based on the results.

[16] also develops mathematical criteria with proofs for local optimality conditions of execution plans of information filters. These conditions can be evaluated in polynomial time, and can be the basis of a greedy algorithm that experimentally demonstrated near-perfect success in finding the globally optimal sequence of a conjunction of fifteen or fewer randomly-generated filters. These conditions derive from a decision-theoretic proeessing-cost model of the the expected cost of sending a data through a conjunctive sequence of filters:

$$t_{1,m} = c_1 + c_2 p(f_1) + c_3 p(f_1 \wedge f_2) + \quad + c_m p(f_1 \wedge f_2 \cdots \wedge f_{m-1})$$

where f_i is the event of passing filter i, $p(f_i)$ is the probability of passing filter i, and c_i is the cost of passing filter i. Then [16] gives a local optimality criterion against interchange of filters i and $i+1$ in the conjunctive filter sequence if:

$$c_i/[1-p(f_i|f_1 \wedge f_2 \wedge \cdots f_{i-1})] \leq c_{i+1}/[1-p(f_{i+1}|f_1 \wedge f_2 \wedge \cdots f_{i-1})]$$

and a local optimality criterion against deletion of redundant but fast filter i:

$$c_i + c_{i+1}p(f_i|f_1 \cdots \wedge f_{i-1}) + c_{i+2}p(f_i \wedge f_{i+1}|f_1 \cdots \wedge f_{i-1}) + \cdots + c_e p(f_i \cdots \wedge f_{e-1}|f_1 \wedge \cdots f_{i-1})$$
$$\leq c_{i+1} + c_{i+2}p(f_{i+1}|f_1 \wedge \cdots f_{i-1}) + \cdots + c_e p(f_{i+1} \cdots \wedge f_{e-1}|f_1 \wedge \cdots f_{i-1})$$

Dual criteria can be proved for disjunctive sequences, on the inverse of the probability involved.

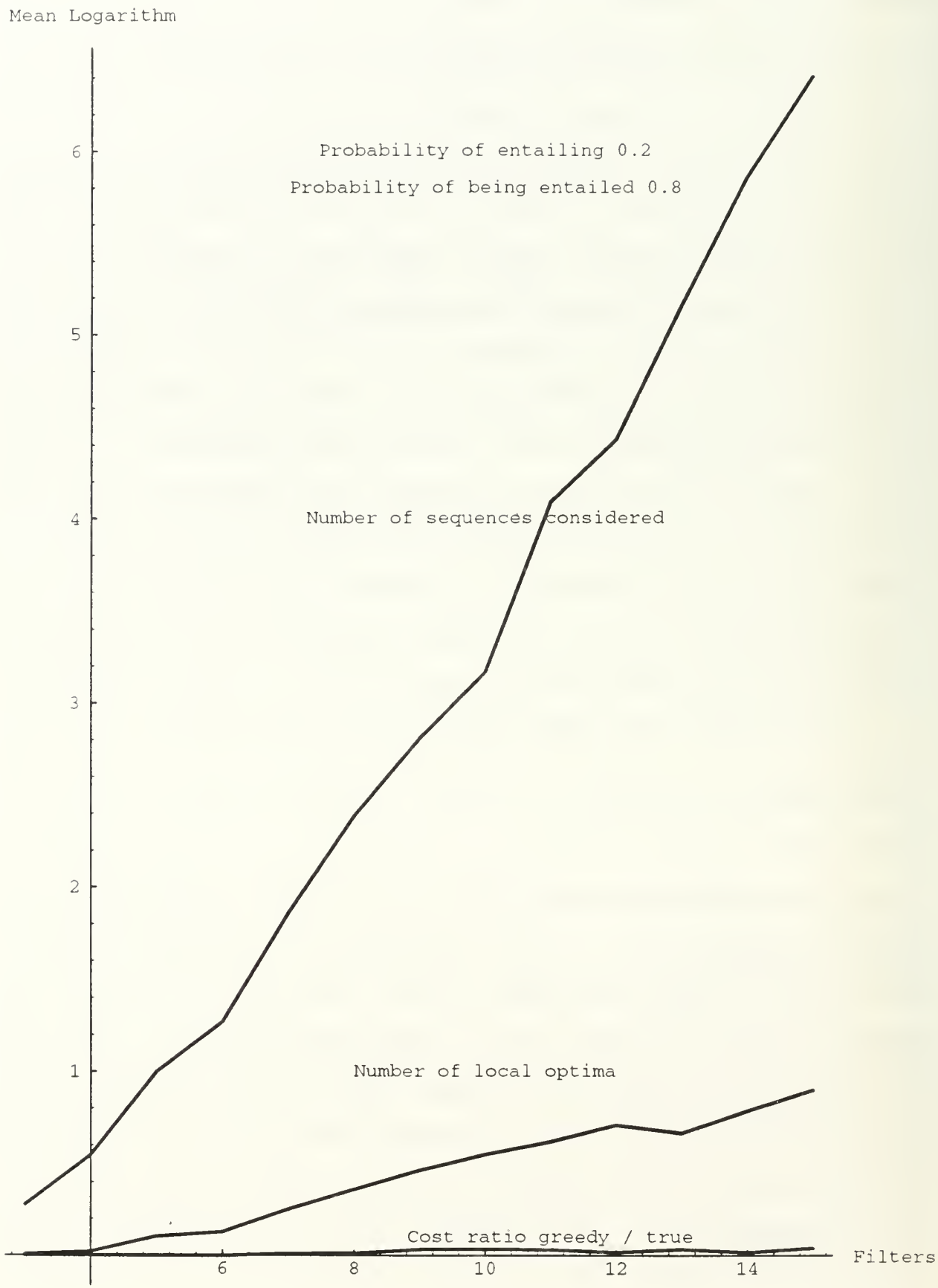
Further local optimality conditions we prove in [16] are that distributive laws should be used to factor terms whenever possible, and that DeMorgan's Laws should be used to push negations in as far as possible in the boolean expression of the sequential filter execution plan. Finally, and most surprisingly, we proved it is never locally optimal to have different information filters operating in parallel, no matter how many additional processors are available, because the increased throughput does not compensate for the increased workload on each filter. This proof makes only broad assumptions: That the cost, per unit number of data items, of n processors doing a filter i is $g(n)+(c_i/n)$, for some g where $g''(n) \leq 0$ and $g(0)=0$. However, using multiple processors on the *same* filter simultaneously is locally optimal under the same processing model, the approach of [22].

The above optimality analysis can be used to find a good consensus execution plan for information filtering for an application, using means of costs and probabilities on a representative set of queries and captions, as we did in [16]. But it can also be used to improve upon the consensus execution plan for a particular query at runtime. If we first apply the consensus execution plan to a small random sample of the input data, we can estimate problem-specific values for costs and probabilities, and replan based on those. This is useful when there are hidden correlations (conditional probabilities) between the words of a query. One application is to deciding whether to interleave index lookups for the particular nouns of the query with other more global analysis of the query. For instance for the query "AIM-9R on an aircraft", "aircraft" is very common in the NAWC-WD captions, and AIM-9Rs are usually shown on aircraft; so the mathematical criteria will say that we ought to first find pictures of AIM-9Rs, then do picture-type matching, and then check to see if the remaining candidate captions mention an aircraft (and then do subgraph matching to confirm that the AIM-9R is on the aircraft and not

beside it).

We confirmed the predictions of our theory of optimal execution plans in two quite different sets of experiments reported in [16]. In one set of experiments we repeatedly generated random filter sets up to size 15, and checked whether a greedy algorithm based on the above local optimality criteria could find the globally optimal conjunctive sequence of those filters. We verified the global optimum by exhaustive search through all possible filter sequences containing the required filters plus some improper subset of the redundant filters. Fig. 2 shows typical results that we obtained, in this case for 13,000 experiments in which costs were even distributed on the range 0 to 10 and probabilities were evenly distributed on the range 0.01 to 1.0. In Fig. 2 in particular, 0.2 was the probability that a filter was redundancy-creating and 0.8 was the probability that a filter was redundant with respect to some redundancy-creating filter later in a conjunctive sequence, parameters close to those of MARIE and the variations on it that we have explored. The horizontal axis is the number of filters considered, and the vertical axis is the mean of the logarithms of the output parameter indicated. It can be seen that the number of local optima grows significantly more slowly than the size of the search space, the number of sequences considered by exhaustive search. The ratio of the cost of the filter sequence found by our polynomial-time greedy algorithm to the cost of the filter sequence found by exhaustive consideration of all possible sequences is very close to unity. Thus even if this problem is exponential in time complexity in the worst case, simple polynomial-time algorithms usually work so well that there is little reason to use anything else with 15 or less filters.

The second set of experiments involved more detailed modeling of MARIE-1, using more detailed parameters derived from 44 test queries, all but 2 of which were supplied by naive users of the existing NAWC-WD system. We estimated cost and probability parameters by running each filter separately on the database of 217 captions used in [17]. We then confirmed that the actual performance of our prototype system on the 44 queries was very close to that predicted by theory. For instance, comparing cost of filters without the picture-type matcher to cost with it, we observed a ratio of 1.18 with a standard deviation of 0.43 versus a predicted ratio of 1.33; and in comparing cost of filters without the keyword matcher to cost with it, we observed a ratio of 22.1 with a standard deviation of 17.3 versus a theoretical ratio of 29.7. In the first comparison, the theoretical


Mean Logarithm
6
5
4
3
2
1
Probability of entailing 0.2
Probability of being entailed 0.8
Number of sequences considered
Number of local optima
Cost ratio greedy / true
6
8
10
12
14
Filters

optimum was optimal in all but 9 of the 44 cases, and in the second comparison, the theoretical optimum was optimal in all 44 cases. These experiments are encouraging. We hope to do further experiments, and explore more filters and more complicated filters.

5. References

[1] Belkin, N. J. and Croft, W. B. Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM, 35*, 12 (December 1992), 29-38

[2] Chang, C.-C. and Wu, T.-C. Retrieving the most similar symbolic pictures from pictorial databases. *Information Processing and Management, 28*, 5, 581-588.

[3] Chang, J. W., Hyuk, J.C., Sang, H. O., and Lee, Y. J. Hybrid access method: an extended two-level signature file approach. International Conference on Multimedia Information Systems, ACM, Singapore (1991), 51-62.

[4] Cochran, W. G. *Sampling Techniques, third edition.* New York: Wiley, 1977.

[5] Constantopoulos, P., Drakopoulos, J., and Yeorgaroudakis, Y. Retrieval of multimedia documents by pictorial content: a prototype system. International Conference on Multimedia Information Systems, ACM, Singapore (1991), 35-48.

[6] Covington, M. *Natural language processing for Prolog programmers.* Englewood Cliffs, NJ: Prentice-Hall, 1994.

[7] Faloutsos, C. Signature-based text retrieval methods: a survey. *Database Engineering*, March 1990, 27-34.

[8] Fujisaki, T., Jelinek, F., Cocke, J., Black, E., and Nishino, T. A probabilistic parsing method for sentence disambiguation. In *Current issues in parsing technology*, ed. Tomita, M., Boston: Kluwer, 1991.

[9] Grosz, B., Appelt, D., Martin, P. and Pereira, F. TEAM: An experiment in the design of transportable natural language interfaces. *Artificial Intelligence, 32* (1987), 173-243.

[10] Jones, M. and Eisner, J. A probablistic parser applied to software testing documents. Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, July 1992, 323-328.

[11] Krovez, R. and Croft, W. B. Lexical ambiguity and information retrieval. *ACM Transactions on Information Systems, 10,* 2 (April 1992), 115-141.

[12] Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. Five papers on Wordnet. *International Journal of Lexicography, 3*, 4 (Winter 1990).

[13] Rau, L. Knowledge organization and access in a conceptual information system. *Information Processing and Management, 23*, 4 (1988), 269-284.

[14] Rowe, N. Antisampling for estimation: an overview. *IEEE Transactions on Software Engineering, SE-11,* 10 (October 1985), 1081-1091.

[15] Rowe, N. Inferring depictions in natural-language captions for efficient access to picture data. To appear in *Information Processing and Management*, 1994.

[16] Rowe, N. Using local optimality criteria for efficient information retrieval with redundant information filters. Technical report, U.S. Naval Postgraduate School, February 1994.

[17] Rowe, N. and Guglielmo, E. Exploiting captions in retrieval of multimedia data. *Information Processing and Management, 29*, 4 (1993), 453-461.

[18] Seem, D. and Rowe, N. Shape correlation of low-frequency underwater sounds. *Journal of the Acoustical Society of America, 90*, 5 (April 1994).

[19] Sembok, T. and van Rijsbergen, C. SILOL: A simple logical-linguistic document retrieval system. *Information Processing and Management, 26*, 1 (1990), 111-134.

[20] Smeaton, A. F. Progress in the application of natural language processing to information retrieval tasks. *The Computer Journal, 35*, 3 (1992), 268-278.

[21] Smith, P., Shute, S., Galdes, D., and Chignell, M. Knowledge-based search tactics for an intelligent intermediary system. *ACM Transactions on Information Systems, 7,* 3 (July 1989), 246-270.

[22] Stanfill, C. and Kahle, B. Parallel free-text search on the Connection Machine system. *Communications of the Assocation for Computing Machinery, 29,* 12 (December 1986), 1229-1239.

Distribution List

Defense Technical Information Center Cameron Station Alexandria, VA 22314	1
Library, Code 52 Naval Postgraduate School Monterey, CA 93943	1
Research Office Code 08 Naval Postgraduate School Monterey, CA 93943	1
Dr.Neil C. Rowe, Code CSRp Naval Postgraduate School Computer Science Department Monterey, CA 93943-5118	50
Mr. Russell Davis HQ, USACDEC Office of Naval Research Attention: ATEC-1M Fort Ord, CA 93941	1
Ralph Wachter, Code 333 Computer Science Office of Naval Research Ballston Tower One 800 North Quincy St. Arlington, VA 22217-5660	1